Thermodynamics of water modeled using *ab initio* simulations

Valéry Weber[*]

Physical Chemistry Institute, University of Zurich, 8057 Zurich, Switzerland

D. Asthagiri[†]

Department of Chemical and Biomolecular Engineering,
Johns Hopkins University, Baltimore, MD 21218
(Dated: August 9, 2010)

We regularize the potential distribution framework to calculate the excess free energy of liquid water simulated with the BLYP-D density functional. The calculated free energy is in fair agreement with experiments but the excess internal energy and hence also the excess entropy are not. Our work emphasizes the importance of thermodynamic characterization in assessing the quality of electron density functionals in describing liquid water and hydration phenomena.

The liquid-vapor coexistence for water is thoroughly characterized experimentally, and the excess free energy (chemical potential) of a water molecule in the liquid relative to the vapor, $\mu_{\mathrm{w}}^{\mathrm{ex}}$, a basic descriptor of the liquid-vapor equilibrium, is very well-established. The chemical potential, together with its derivatives, especially the temperature derivative (the excess entropy), are essential quantities in understanding hydration phenomena and chemical transformations in the liquid state.

For *ab initio* simulations of water, two earlier studies [1, 2] have sought $\mu_{\mathrm{w}}^{\mathrm{ex}}$. The first study revealed how the overbinding of water by PW91 (PBE) functionals leads to an overly negative chemical potential and ultimately an over-structured fluid. The second explored the coexisting densities of the liquid and vapor phases [2] and found a higher than normal vapor density, indicating overbinding of molecules by the BLYP functional. Their limitations notwithstanding, these early studies were incisive in evaluating the fluid simulated by *ab initio* dynamics.

In this Letter we present a rigorous calculation of the excess free energy using a theoretical approach that has matured over the last few years [3–5], one that renders calculating free energies from *ab initio* simulations far less daunting than before. Together with an independent calculation of the excess internal energy, we obtain the excess entropy as well. A key finding of this work is that a satisfactory agreement with experiments of the chemical potential can mask errors in the excess energy and the entropy, quantities that are crucial in understanding the thermodynamics of hydration.

The relation between $\mu_{\mathrm{w}}^{\mathrm{ex}}$ and intermolecular interactions is given by the potential distribution theorem [6, 7]

$$\beta\mu_{\mathrm{w}}^{\mathrm{ex}} = \ln \int e^{\beta\varepsilon} P(\varepsilon) d\varepsilon \,, \qquad (1)$$

where $\varepsilon = U_N - U_{N-1} - U_{\mathrm{w}}$ is the binding energy of the distinguished water molecule with the rest of the fluid. U_N is the potential energy of the N-particle system at a particular configuration, U_{N-1} is the potential energy of the configuration but with the distinguished water removed, and U_{w} is the potential energy of the distinguished water molecule solely. $P(\varepsilon)$ is the probability density distribution of ε and is obtained by sampling many configurations of the system. $\mu_{\mathrm{w}}^{\mathrm{ex}}$ is the excess free energy in the liquid relative to an ideal gas at the same density and temperature. Based on the experimental coexistence densities [8] at 298 K, $\mu_{\mathrm{w}}^{\mathrm{ex}} = -6.3$ kcal/mol.

A naive application of Eq. 1 to liquid water will fail because the high energy regions of $P(\varepsilon)$, reflecting the short-range repulsive interactions [4, 9, 10], are never well sampled in a simulation. We resolve this difficulty by regularizing Eq. 1 [10]. Consider a hard-core solute of radius λ centered on a distinguished water molecule: the hard-core solute excludes the remaining water oxygen atoms from within the sphere of radius λ. The chemical potential of the hard-core solute, $\mu_{\mathrm{HC}}^{\mathrm{ex}}$, is assumed known. With this construction, Eq. 1 can be rewritten as [4, 9, 10]:

$$\beta\mu_{\mathrm{w}}^{\mathrm{ex}} = \beta\mu_{\mathrm{HC}}^{\mathrm{ex}} + \ln x_0 + \ln \int P(\varepsilon|n_\lambda = 0) e^{\beta\varepsilon} d\varepsilon \,. \quad (2)$$

Here $P(\varepsilon|n_\lambda = 0)$ is the binding energy distribution of the distinguished water molecule conditioned on there being no ($n_\lambda = 0$) water oxygen atoms within λ of the distinguished oxygen atom. By moving the boundary away from the distinguished water, we temper the interaction of the distinguished water with the rest of the fluid; indeed, for a sufficiently large λ, $P(\varepsilon|n_\lambda = 0)$ is expected to be well-described by a Gaussian [4]. The fraction of configurations that do not have any oxygen atoms within λ of the distinguished oxygen is x_0; this is the $n = 0$ member of the set $\{x_n\}$ of coordination states sampled by the distinguished water molecule and characterizes the interactions between the distinguished water and the contents of the inner-shell. The excess chemical potential of the hard-core solute $\beta\mu_{\mathrm{HC}}^{\mathrm{ex}} = -\ln p_0$, where p_0 is the fraction of configurations in which a cavity of radius λ is found in the liquid. p_0 is the $n = 0$ member of the set $\{p_n\}$ of occupation numbers of a cavity and characterizes the packing interactions in hydration [11].

Since x_0 and p_0 are directly obtained from the simulation record, we only need explicit energy evaluations to

compute the outer-shell contribution

$$\beta\mu_{\text{outer}}^{\text{ex}} = \ln \int P(\varepsilon|n_\lambda = 0)e^{\beta\varepsilon}d\varepsilon. \qquad (3)$$

Here we calculate the outer-shell contribution using an alternative expression

$$\beta\mu_{\text{outer}}^{\text{ex}} = -\ln \int P^{(0)}(\varepsilon|n_\lambda = 0)e^{-\beta\varepsilon}d\varepsilon, \qquad (4)$$

where $P^{(0)}(\varepsilon|n_\lambda = 0)$ is the binding energy distribution in the uncoupled ensemble [6, 10]; that is, after we find a suitable cavity in the liquid, we insert a test-particle in that cavity and assess $P^{(0)}(\varepsilon|n_\lambda = 0)$. The superscript (0) indicates that the test-particle and the fluid are thermally uncoupled. It is advantageous to use Eq. 4 over Eq. 3 because the uncoupled distribution $P^{(0)}(\varepsilon|n_\lambda = 0)$ has a higher entropy than the coupled distribution $P(\varepsilon|n_\lambda = 0)$, rendering the calculation of the free energy of inserting the particle (Eq. 4) more robust than extracting it (Eq. 3) [12]. Further, when x_0 is small, it is difficult to characterize $P(\varepsilon|n_\lambda = 0)$ reliably, a limitation that does not apply to $P^{(0)}(\varepsilon|n_\lambda = 0)$.

For λ sufficiently large such that $P^{(0)}(\varepsilon|n_\lambda = 0)$ is Gaussian distributed with a mean $\langle\varepsilon|n_\lambda = 0\rangle_0$ and variance $\langle\delta\varepsilon^2|n_\lambda = 0\rangle_0$ (the subscript 0 emphasizes that the test-particle and the fluid are thermally uncoupled), Eq. 4 becomes

$$\beta\mu_{\text{outer}}^{\text{ex}} = \langle\varepsilon|n_\lambda = 0\rangle_0 - \frac{1}{2k_\text{B}T}\langle\delta\varepsilon^2|n_\lambda = 0\rangle_0. \qquad (5)$$

We apply the above framework to water simulated with the BLYP-D electron density functional. (In BLYP-D, the BLYP [13, 14] functional is augmented with empirical corrections for dispersion interactions [15].) We simulate the liquid at a density of 0.997 g/cm³ (number density of 33.33 nm⁻³) and temperature of 350 K. The higher temperature effectively weakens the bonding [5] and has been found necessary to model the real liquid at the standard density and $T = 298$ K [5, 16].

We simulate water with the CP2K code[17] and in the NVE and NVT ensembles; the electronic structure calculations are exactly as in our earlier study [5], and only salient differences are noted here. For the studies at NVT, we use the hybrid Monte Carlo method and extended the simulations in our earlier study of a 32 water system [5]. In the NVE ensemble, we simulated systems with 32 and 64 water molecules.

The production phase of the hybrid Monte Carlo lasted 2585 sweeps ($\approx$130 ps). The system temperature was 350 K. Simulations in the NVE ensemble lasted 200 ps of which the last 170 ps were used for analysis. The initial configuration for the 32 particle system was obtained from the end-point of our earlier study [5]. The initial configuration for the 64 water system was obtained from an equilibrated configuration of SPC/E water [18]

molecules. In the production phase, the average temperature was 357 ± 26 K (362 ± 19 K) for the 32 (64) water system. The energy drift was less than 1.8 K (1.2 K) for the 32 (64) water system.

To calculate $\mu_{\text{outer}}^{\text{ex}}$, we construct a Cartesian grid within the simulation cell. The oxygen population within a radius λ of each grid center is calculated. In instances where there are no ($n_\lambda = 0$) water molecules, we insert a test water molecule in the cavity, assess its binding energy, and thus construct $P^{(0)}(\varepsilon|n_\lambda = 0)$. (We draw water molecules from the simulation cell, randomly rotate it about an axis through the oxygen center, and use the resulting configuration as the test particle.)

In the NVE simulations, we first find approximately 3500 cavities from configurations sampled every 50 fs. (The target count of cavities is met by adjusting the spacing of the Cartesian grid between approximately 0.3 to 0.96 Å.) Thus over 110k($\approx 32 * 3500$) binding energy calculations are used to construct $P^{(0)}(\varepsilon|n_\lambda = 0)$. In the NVT simulations, the grid spacing for finding cavities to compute $P^{(0)}(\varepsilon|n_\lambda = 0)$ was fixed at 2.0 Å. For $\lambda = 3.0$ Å this gave 98 cavities. In this instance, five different orientations were used for each test particle giving about 16k($\approx 5 * 32 * 98$) binding energies to construct $P^{(0)}(\varepsilon|n_\lambda = 0)$. Subsequently, we resampled the trajectory using a finer grid to better estimate the probability of finding p_0. Throughout, uncertainties in x_0, p_0, and $\langle e^{-\beta\varepsilon}|n_\lambda = 0\rangle$ were estimated using the Friedberg and Cameron [19] block transformation procedure [20].

Figure 1 shows $P^{(0)}(\varepsilon|n_\lambda = 0)$ for different cavity radii for the 32 water system. As the figures shows, the low energy region is well-characterized for inner-shell radii considered here; and it is this low-ε part of the distribution that is most important for calculating $\mu_{\text{outer}}^{\text{ex}}$ (Eq. 4).

Figure 1 shows that the agreement between $P^{(0)}(\varepsilon|n_\lambda = 0)$ obtained in the NVE and NVT simulations is satisfactory, although there is somewhat more scatter in the low-energy region for results with NVT. This is a consequence of using an order of magnitude less data (16k versus 110k) to construct $P^{(0)}(\varepsilon|n_\lambda = 0)$.

The probability x_0 is a direct measure of the solute-water interactions within the inner-shell, and perhaps not surprisingly, for large λ, this quantity is very low. In this case, one can model $\{x_n\}$ with a maximum entropy approach [21, 22] using the robust estimates of mean and variance of $\{x_n\}$ and thus secure x_0. But for large λ [3, 5], the underlying assumption of Gaussian occupancy statistics may not be valid. Here, we instead use Bayes's theorem in the form [3]

$$x_n = p_{n+1}\frac{p(1|n+1)}{\sum_{m\geq 1} p_m p(1|m)}, \qquad (6)$$

where $p(1|m)$ is the conditional probability of finding one water molecule at the center of the cavity given that m
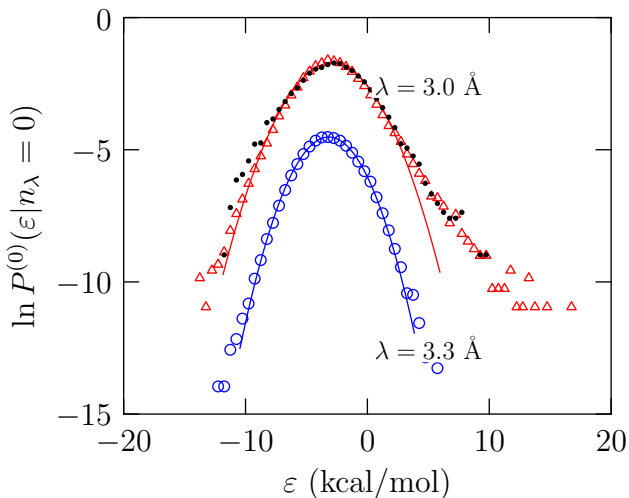
FIG. 1. Distribution of interaction energies $P^{(0)}(\varepsilon|n_\lambda = 0)$ for a distinguished water molecule in the center of an empty cavity of radius λ. The system comprises 32 water molecules. The red triangles and blue circles are from the NVE simulation. The solid line defines the Gaussian model for the respective distribution. The black dots are from the NVT simulation. The $\lambda = 3.3$ Å curve has been shifted downwards by 3 units for clarity. Observe that the thermodynamically important (Eq. 4) low-energy tail is well-characterized. Further, for $\lambda = 3.3$ Å, both wings of the distribution are tempered and the Gaussian model becomes a good approximation.

water molecules are present in the cavity. For this purpose, following an earlier study [3], the center is any point within 0.15 Å of the geometric center of the cavity.

Figure 2 depicts $\{x_n\}$ for the 64 particles system. The fit using Eq. 6 is in excellent agreement with the actual data. This gives us confidence in the estimated x_0 for $\lambda = 3.3$ Å (Fig. 2), an instance where x_0 was not observed in the simulation. As Fig. 2 shows, the maximum entropy model can lead to errors in $k_B T \ln x_0$ by about a kcal/mol or more, especially for the large λ.

Table I collects the various components of the excess chemical potential. Across the range of λ, the calculated values are within a kcal/mol (often less) of each other and the experimental value of -6.3 kcal/mol for liquid water at 298 K. Comparing the 32 and 64 water results shows modest system size effect of about $k_B T$. This difference primarily arises from a less favorable μ_{outer}^{ex} for the larger system: for a large cavity in a small simulation cell, the outer-shell water molecules pack tightly at the interface and lead to the lower μ_{outer}^{ex}.

We can calculate the excess entropy of hydration per particle from excess chemical potential and the mean binding energy, $\langle \varepsilon \rangle$, of a distinguished water molecule in the fully coupled simulation. First consider the excess internal energy. For the 32 (64) water system $\langle \varepsilon \rangle = -28$ kcal/mol (-27.3 kcal/mol). Thus the heat of vaporization per particle $\Delta H_{vap} \approx -\langle \varepsilon \rangle/2 + k_B T =$
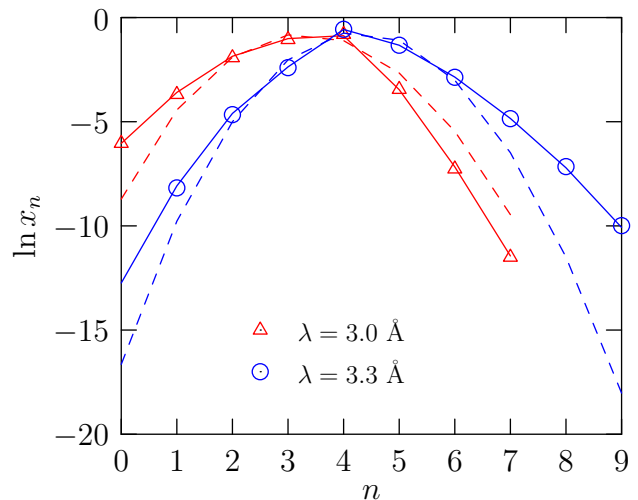

FIG. 2. Observed and estimated $\{x_n\}$ for a $N = 64$ water molecule system and different inner-shell radii λ. Symbols denote the observed data. The solid line is obtained using Bayes' theorem (Eq. 6). The dashed lines are maximum entropy fits using the mean and variance of $\{x_n\}$ and a flat prior [22]. Observe that $\{x_n\}$ obtained using Eq. 6 well-describes the simulation results.

14.6 kcal/mol (14.2 kcal/mol) is substantially in error relative to the experimental value of about 10.5 kcal/mol at 300 K [8].

Neglecting small contributions due to the thermal expansion and compressibility of the medium, the excess entropy per particle [4, 10] is $S^{ex}/nk_B \approx \langle \varepsilon \rangle/2k_B T - \mu^{ex}/k_B T$. Based on the estimates of $\langle \varepsilon \rangle$ and μ_w^{ex} above, it is clear that TS^{ex} for BLYP-D water is also in error.

Several important lessons emerge from the present study. First, while BLYP-D at 350 K appears to adequately describe the structure [5] and the hydration free energy relative to experiments, the agreement comes due to balancing errors in the excess entropy and excess internal energy. Second, while dispersion interactions are regarded as necessary in describing water, the empirical dispersion correction in BLYP-D overbinds the material and the temperature of 350 K does not adequately compensate for this overbinding. Finally, the BLYP-D functional at 350 K should not be expected to describe the hydration thermodynamics of aqueous solutes, especially properties such as the excess entropy of hydration, and thus also transport properties of solutes in the medium.

Acknowledgments

TABLE I. Chemical ($k_\mathrm{B}T \ln x_0$), packing ($k_\mathrm{B}T \ln p_0$), and long-range interaction (μ_{outer}^{ex}) contributions to the excess chemical potential (μ_w^{ex}) of water obtained from NVE molecular dynamics simulations and NVT hybrid Monte Carlo simulations [5]. The radius of the inner shell is λ (in Å). The average temperatures of the $N = 32$ and $N = 64$ water systems are, respectively, 357 ± 26 K and 362 ± 19 K. The temperature is 350 K for simulation in the NVT ensemble. Where available, the inner-shell chemical contribution is evaluated directly from the data, else the estimate using Eq. 6 is used. The outer-shell contribution is calculated by numerical integration of Eq. 4; estimates from a Gaussian model 5 (data in parenthesis) are provided for comparison. Energies are in kcal/mol. Uncertainties are at 1σ level.

N	λ	$k_\mathrm{B}T \ln x_0$ (Simulation)	$k_\mathrm{B}T \ln x_0$ (Bayes)	$-k_\mathrm{B}T \ln p_0$	μ_{outer}^{ex}	$\mu_\mathrm{w}^{ex}(l)$
32 (NVE)	3.0	-4.3 ± 0.1	-4.3	5.5 ± 0.1	$-7.1 \pm 0.2 \ (-6.5)$	-5.9 ± 0.2
	3.1	-5.6 ± 0.1	-5.5	6.2 ± 0.2	$-6.5 \pm 0.2 \ (-6.0)$	-5.9 ± 0.3
	3.2	-7.2 ± 0.4	-6.8	7.0 ± 0.2	$-6.2 \pm 0.2 \ (-5.8)$	-6.4 ± 0.5
	3.3	-8.5 ± 0.6	-8.6	8.0 ± 0.3	$-5.7 \pm 0.1 \ (-5.5)$	-6.2 ± 0.7
(NVT)	3.0	-4.4 ± 0.1	-4.3	5.2 ± 0.3	$-7.0 \pm 0.2(-7.2)$	-6.2 ± 0.4
64 (NVE)	3.0	-4.4 ± 0.0	-4.3	5.5 ± 0.2	$-6.8 \pm 0.5 \ (-5.6)$	-5.7 ± 0.5
	3.1	-5.8 ± 0.1	-5.7	6.1 ± 0.2	$-5.8 \pm 0.3 \ (-5.2)$	-5.5 ± 0.4
	3.2	-7.2 ± 0.2	-7.2	6.7 ± 0.3	$-5.3 \pm 0.1 \ (-4.7)$	-5.8 ± 0.4
	3.3	not observed	-9.2	7.3 ± 0.3	$-4.8 \pm 0.1 \ (-4.2)$	-6.7

* Email: valeryweber@hotmail.com
† Email: dilipa@jhu.edu

[1] D. Asthagiri, L. R. Pratt, and J. D. Kress, Phys. Rev. E **68**, 041505 (2003).
[2] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, M. Sprik, J. Hutter, F. Mohammed, M. Krack, and M. Parrinello, J. Phys. Chem. A **110**, 640 (2006).
[3] A. Paliwal, D. Asthagiri, L. R. Pratt, H. S. Ashbaugh, and M. E. Paulaitis, J. Chem. Phys. **124**, 224502 (2006).
[4] J. K. Shah, D. Asthagiri, L. R. Pratt, and M. E. Paulaitis, J. Chem. Phys. **127**, 144508 (2007).
[5] V. Weber, S. Merchant, P. D. Dixit, and D. Asthagiri, J. Chem. Phys. **132**, 204509 (2010).
[6] T. L. Beck, M. E. Paulaitis, and L. R. Pratt, *The potential distribution theorem and models of molecular solutions* (Cambridge University Press, 2006).
[7] B. Widom, J Phys Chem **86**, 869 (1982).
[8] W. Wagner and A. Pruss, J. Phys. Chem. Ref. Data **31**, 387 (2002).
[9] D. Asthagiri, H. S. Ashbaugh, A. Piryatinski, M. E. Paulaitis, and L. R. Pratt, J. Am. Chem. Soc. **129**, 10133 (2007).
[10] D. Asthagiri, S. Merchant, and L. R. Pratt, J. Chem. Phys. **128**, 244512 (2008).
[11] L. R. Pratt, Ann. Rev. Phys. Chem. **53**, 409 (2002).
[12] N. Lu and D. A. Kofke, J. Chem. Phys. **114**, 7303 (2001).
[13] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[14] C. T. Lee, W. T. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[15] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[16] J. Schmidt, J. VandeVondele, I. F. W. Kuo, D. Sebastini, J. I. Siepmann, J. Hutter, and C. J. Mundy, J. Phys. Chem. B **113**, 11959 (2009).
[17] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comp. Phys. Comm. **167**, 103 (2005).
[18] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[19] R. Friedberg and J. E. Cameron, J. Chem. Phys. **52**, 6049 (1970).
[20] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 1987), chap. 6. How to analyze the results, pp. 192–195.
[21] G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt, Proc. Natl. Acad. Sc. USA **93**, 8951 (1996).
[22] G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, J. Phys. Chem. B **102**, 10469 (1998).